



Sequence Listing

<110> Harms, Jerome S.
Splitter, Gary A.
Eakle, Kurt A.
Bremel, Robert D.

<120> Inducible Protein Expression System

<140> US 10/763,976

<141> 2004-01-23

<160> 13

<210> 1

<211> 576

<212> DNA

<213> Artificial Sequence

<220>

<221> Promoter

<222> 87..432

<223> BLV Promoter

<220>

<221> misc_feature

<222> 452..576

<223> attR1 Gateway recombination site

<400> 1

AGGAAACCAG CAGCGGCTAT CCGCGCATCC ATGCCCCCGA ACTGCAGGAG TGGGGAGGCA 60
CGATGGCCGC TTTGGTCGAG GCGGATCCTA GCAGAAAAAT AAGACTTGAT TCCCCCTTAA 120
AATTACAAC TCTAGAAAAT GAATGGCTCT CCCGCCTTTT TTGAGGGGGA ATCATTGTGA 180
TGAAAGATCA TGCCGACCTA GGCGCCGCCA CCGCCCCGTA AACCAGACAG AGACGTCAGC 240
TGCCAGAAAA GCTGGTGACG GCAGCTGGTG GCTAGAATCC CCGTACCTCC CCAACTTCCC 300
CTTTCCCGAA AAATCCACAC CCTGAGCTGC TGACCTCACC TGCTGATAAA TTAATAAAAT 360
GCCGGCCCTG TCGAGTTAGC GGCACCAGAA GCGTTCTTCT CCTGAGACCC TCGTGCTCAG 420
CTCTCGGTCC TGCCTCGAGA AGCTTGTTAT CACAAGTTTG TACAAAAAAG CTGAACGAGA 480
AACGTAAAAT GATATAAATA TCAATATATT AAATTAGATT TTGCATAAAA AACAGACTAC 540
ATAATACTGT AAAACACAAC ATATCCAGTC ACTATG 576

<210> 2
 <211> 930
 <212> DNA
 <213> bovine leukemia virus

<220>
 <221> CDS
 <222> 1..930

<400> 2

```

ATG GCA AGT GTT GTT GGT TGG GGG CCC CAC TCT CTA CAT GCC TGC 45
Met Ala Ser Val Val Gly Trp Gly Pro His Ser Leu His Ala Cys
1          5          10          15

CCG GCC CTG GTT TTG TCC AAT GAC GTC ACC ATC GAT GCC TGG TGC 90
Pro Ala Leu Val Leu Ser Asn Asp Val Thr Ile Asp Ala Trp Cys
          20          25          30

CCC CTC TGC GGG CCC CAT GAG CGA CTC CAA TTC GAA AGG ATC GAC 135
Pro Leu Cys Gly Pro His Glu Arg Leu Gln Phe Glu Arg Ile Asp
          35          40          45

ACC ACG CAC ACC TGC GAG ACC CAC CGT ATC ACC TGG ACC GCC GAT 180
Thr Thr His Thr Cys Glu Thr His Arg Ile Thr Trp Thr Ala Asp
          50          55          60

GGA CGA CCT TTC GGC CTC AAT GGA GCG CTG TTC CCT CGA CTG CAT 225
Gly Arg Pro Phe Gly Leu Asn Gly Ala Leu Phe Pro Arg Leu His
          65          70          75

GTC TCC AGA GAC CCG GCC CCA AGG GCC CGA CGA CTC TGG ATC AAC 270
Val Ser Arg Asp Pro Ala Pro Arg Ala Arg Arg Leu Trp Ile Asn
          80          85          90

TGC CCC CTT CCG GCC GTT CGC GCT CAG CCC GGC CCG GTT TCA CTT 315
Cys Pro Leu Pro Ala Val Arg Ala Gln Pro Gly Pro Val Ser Leu
          95          100          105

TCC CCC TTC GAG CGG TCC CCC TTC CAG CCC TAC CAA TGC CAA TTG 360
Ser Pro Phe Glu Arg Ser Pro Phe Gln Pro Tyr Gln Cys Gln Leu
          110          115          120

CCC TCG GCC TCT AGC GAC GGT TGC CCC GTC ATC GGG CAC GGC CTT 405
Pro Ser Ala Ser Ser Asp Gly Cys Pro Val Ile Gly His Gly Leu
          125          130          135

CTT CCC TGG AAC AAC TTA GTA ACG CAT CCT TGT CCT CGG AAA GTC 450
Leu Pro Trp Asn Asn Leu Val Thr His Pro Cys Pro Arg Lys Val
          140          145          150

CTT ATA TTA AAT CAA ATG GCC AAT TTT TCC TTA CTC CCC CCC TTC 495
Leu Ile Leu Asn Gln Met Ala Asn Phe Ser Leu Leu Pro Pro Phe
          155          160          165

AAT ACC CTC CTT GTG GAC CCC CTC CGG TTG TCC GTC TTT GCC CCA 540
Asn Thr Leu Leu Val Asp Pro Leu Arg Leu Ser Val Phe Ala Pro

```

	170	175	180	
GAC ACC AGG GGA GCC ATA CGT TAT CTC TCC ACC CTT TTG ACG CTA	585			
Asp Thr Arg Gly Ala Ile Arg Tyr Leu Ser Thr Leu Leu Thr Leu				
	185	190	195	
TGC CCA GCT ACT TGT ATT CTA CCC CTC GGC GAG CCC TTC TCT CCT	630			
Cys Pro Ala Thr Cys Ile Leu Pro Leu Gly Glu Pro Phe Ser Pro				
	200	205	210	
AAT GTC CCC ATA TGT CGC TTT CCC CGG GAC TCC AAT GAA CCC CCC	675			
Asn Val Pro Ile Cys Arg Phe Pro Arg Asp Ser Asn Glu Pro Pro				
	215	220	225	
CTT TCA GAA TTC GAG CTG CCC CTT ATC CAA ACG CCC GGC CTG TCT	720			
Leu Ser Glu Phe Glu Leu Pro Leu Ile Gln Thr Pro Gly Leu Ser				
	230	235	240	
TGG TCT GTC CCC GCG ATC GAC CTA TTC CTA ACC GGC CCC CCT TCC	765			
Trp Ser Val Pro Ala Ile Asp Leu Phe Leu Thr Gly Pro Pro Ser				
	245	250	255	
CCA TGC GAC CGG TTA CAC GTA TGG TCC AGT CCT CAG GCC TTA CAG	810			
Pro Cys Asp Arg Leu His Val Trp Ser Ser Pro Gln Ala Leu Gln				
	260	265	270	
CGC TTC CTC CAT GAC CCT ACG CTA ACC TGG TCA GAA TTG GTT GCT	855			
Arg Phe Leu His Asp Pro Thr Leu Thr Trp Ser Glu Leu Val Ala				
	275	280	285	
AGC AGG AAA CTA AGA CTT GAT TCA CCC TTA AAA TTA CAA CTG TTA	900			
Ser Arg Lys Leu Arg Leu Asp Ser Pro Leu Lys Leu Gln Leu Leu				
	290	295	300	
GAA AAT GAA TGG CTC TCC CGC CTT TTT TGA			930	
Glu Asn Glu Trp Leu Ser Arg Leu Phe ***				
	305			

<210> 3
<211> 1062
<212> DNA
<213> human T-lymphotropic virus 1

<220>
<221> CDS
<222> 1..1059

<400> 3

```
ATG GCC CAC TTC CCA GGG TTT GGA CAG AGT CTT CTT TTC GGA TAC 45
Met Ala His Phe Pro Gly Phe Gly Gln Ser Leu Leu Phe Gly Tyr
1          5          10          15

CCA GTC TAC GTG TTT GGA GAC TGT GTA CAA GGC GAC TGG TGC CCC 90
Pro Val Tyr Val Phe Gly Asp Cys Val Gln Gly Asp Trp Cys Pro
          20          25          30

ATC TCT GGG GGA CTA TGT TCG GCC CGC CTA CAT CGT CAC GCC CTA 135
Ile Ser Gly Gly Leu Cys Ser Ala Arg Leu His Arg His Ala Leu
          35          40          45

CTG GCC ACC TGT CCA GAG CAT CAG ATC ACC TGG GAC CCC ATC GAT 180
Leu Ala Thr Cys Pro Glu His Gln Ile Thr Trp Asp Pro Ile Asp
          50          55          60

GGA CGC GTT ATC GGC TCA GCT CTA CAG TTC CTT ATC CCT CGA CTC 225
Gly Arg Val Ile Gly Ser Ala Leu Gln Phe Leu Ile Pro Arg Leu
          65          70          75

CCC TCC TTC CCC ACC CAG AGA ACC TCT AAG ACC CTC AAG GTC CTT 270
Pro Ser Phe Pro Thr Gln Arg Thr Ser Lys Thr Leu Lys Val Leu
          80          85          90

ACC CCG CCA ATC ACT CAT ACA ACC CCC AAC ATT CCA CCC TCC TTC 315
Thr Pro Pro Ile Thr His Thr Thr Pro Asn Ile Pro Pro Ser Phe
          95          100          105

CTC CAG GCC ATG CGC AAA TAC TCC CCC TTC CGA AAT GGA TAC ATG 360
Leu Gln Ala Met Arg Lys Tyr Ser Pro Phe Arg Asn Gly Tyr Met
          110          115          120

GAA CCC ACC CTT GGG CAG CAC CTC CCA ACC CTG TCT TTT CCA GAC 405
Glu Pro Thr Leu Gly Gln His Leu Pro Thr Leu Ser Phe Pro Asp
          125          130          135

CCC GGA CTC CGG CCC CAA AAC CTG TAC ACC CTC TGG GGA GGC TCC 450
Pro Gly Leu Arg Pro Gln Asn Leu Tyr Thr Leu Trp Gly Gly Ser
          140          145          150

GTT GTC TGC ATG TAC CTC TAC CAG CTT TCC CCC CCC ATC ACC TGG 495
Val Val Cys Met Tyr Leu Tyr Gln Leu Ser Pro Pro Ile Thr Trp
          155          160          165

CCC CTC CTG CCC CAC GTG ATT TTT TGC CAC CCC GGC CAG CTC GGG 540
Pro Leu Leu Pro His Val Ile Phe Cys His Pro Gly Gln Leu Gly
```

170	175	180
GCC TTC CTC ACC AAT GTT CCC TAC AAG CGA ATA GAA GAA CTC CTC 585		
Ala Phe Leu Thr Asn Val Pro Tyr Lys Arg Ile Glu Glu Leu Leu		
185	190	195
TAT AAA ATT TCC CTT ACC ACA GGG GCC CTA ATA ATT CTA CCC GAA 630		
Tyr Lys Ile Ser Leu Thr Thr Gly Ala Leu Ile Ile Leu Pro Glu		
200	205	210
GAC TGT TTG CCC ACC ACC CTT TTC CAG CCT GTT AGG GCA CCC GTC 675		
Asp Cys Leu Pro Thr Thr Leu Phe Gln Pro Val Arg Ala Pro Val		
215	220	225
ACG CTA ACA GCC TGG CAA AAC GGC CTC CTT CCG TTC CAC TCA ACC 720		
Thr Leu Thr Ala Trp Gln Asn Gly Leu Leu Pro Phe His Ser Thr		
230	235	240
CTC ACC ACT CCA GGC CTT ATT TGG ACA TTT ACC GAT GGC ACG CCT 765		
Leu Thr Thr Pro Gly Leu Ile Trp Thr Phe Thr Asp Gly Thr Pro		
245	250	255
ATG ATT TCC GGG CCC TGC CCT AAA GAT GGC CAG CCA TCT TTA GTA 810		
Met Ile Ser Gly Pro Cys Pro Lys Asp Gly Gln Pro Ser Leu Val		
260	265	270
CTA CAG TCC TCC TCC TTT ATA TTT CAC AAA TTT CAA ACC AAG GCC 855		
Leu Gln Ser Ser Ser Phe Ile Phe His Lys Phe Gln Thr Lys Ala		
275	280	285
TAC CAC CCC TCA TTT CTA CTC TCA CAC GGC CTC ATA CAG TAC TCT 900		
Tyr His Pro Ser Phe Leu Leu Ser His Gly Leu Ile Gln Tyr Ser		
290	295	300
TCC TTT CAT AAT TTA CAT CTC CTG TTT GAA GAA TAC ACC AAC ATC 945		
Ser Phe His Asn Leu His Leu Leu Phe Glu Glu Tyr Thr Asn Ile		
305	310	315
CCC ATT TCT CTA CTT TTT AAC AAA AAA GAG GCA GAT GAC AAT GAC 990		
Pro Ile Ser Leu Leu Phe Asn Lys Lys Glu Ala Asp Asp Asn Asp		
320	325	330
CAT GAG CCC CAA ATA TCC CCC GGG GGC TTA GAG CCT CCC AGT GAA 1035		
His Glu Pro Gln Ile Ser Pro Gly Gly Leu Glu Pro Pro Ser Glu		
335	340	345
AAA CAT TTC CGC GAA ACA GAA GTC TGA		1062
Lys His Phe Arg Glu Thr Glu Val ***		
350		

<210> 4
<211> 353
<212> DNA
<213> human T-lymphotropic virus 1

<220>
<221> promoter
<222> 1..353

<400> 4

```
TGACAATGAC CATGAGCCCC AAATATCCCC CGGGGGCTTA GAGCCTCTCA GTGAAAAACA 60
TTTCCGTGAA ACAGAAGTCT GAGAAGGTCA GGGCCCAGAA TAAGGCTCTG ACGTCTCCCC 120
CCGGAGGACA GCTCAGCACC AGCTCAGGCT AGGCCCTGAC GTGTCCCCCT AAAGACAAAT 180
CATAAGCTCA GACCTCCGGG AAGCCACCGG GAACCACCCA TTTCTTCCCC ATGTTTGTCA 240
AGCCGTCCTC AGGCGTTGAC GACAACCCCT CACCTCAAAA AACTTTTCAT GGCACGCATA 300
CGGCTCAATA AAATAACAGG AGTCTATAAA AGCGTGGGGA CAGTTCAGGA GGG          353
```

<210> 5
<211> 456
<212> DNA
<213> human immunodeficiency virus 1

<220>
<221> Promoter
<222> 1..456

<400> 5

```
CTGGAAGGGC TAATTTGGTC CCAAAGAAGA CAAGAGATCC TTGATCTGTG GATCTACCAC 60
ACACAAGGCT ACTTCCCTGA TTGGCAGAAT TACACACCAG GGCCAGGGAT CAGATATCCA 120
CTGACCTTTG GATGGTGCTT CAAGCTAGTA CCAGTTGAGC CAGAGAAGGT AGAAGAGGCC 180
AATGAAGGAG AGAACAACAG CTTGTTACAC CCTATGAGCC TGCATGGGAT GGAGGACGCG 240
GAGAAAGAAG TGTTAGTGTG GAGGTTTGAC AGCAAAC TAG CATTTCATCA CATGGCCCGA 300
GAGCTGCATC CGGAGTACTA CAAAGACTGC TGACATCGAG CTTTCTACAA GGGACTTTCC 360
GCTGGGGACT TTCCAGGGAG GCGTGGCCTG GCGGGGACTG GGGAGTGGCG TCCCTCAGAT 420
GCTGCATATA AGCAGCTGCT TTTTGCCTGT ACTGGG                                456
```

<210> 6
 <211> 306
 <212> DNA
 <213> human immunodeficiency virus 1

<220>
 <221> CDS
 <222> 1..303

<400> 6

ATG	GAG	CCA	GTA	GAT	CCT	AAT	CTA	GAG	CCC	TGG	AAG	CAT	CCA	GGA	45
Met	Glu	Pro	Val	Asp	Pro	Asn	Leu	Glu	Pro	Trp	Lys	His	Pro	Gly	
1				5					10					15	
AGT	CAG	CCT	AGG	ACT	GCT	TGT	AAC	AAT	TGC	TAT	TGT	AAA	AAG	TGT	90
Ser	Gln	Pro	Arg	Thr	Ala	Cys	Asn	Asn	Cys	Tyr	Cys	Lys	Lys	Cys	
				20					25					30	
TGC	TTT	CAT	TGC	TAC	GCG	TGT	TTC	ACA	AGA	AAA	GGC	TTA	GGC	ATC	135
Cys	Phe	His	Cys	Tyr	Ala	Cys	Phe	Thr	Arg	Lys	Gly	Leu	Gly	Ile	
				35					40					45	
TCC	TAT	GGC	AGG	AAG	AAG	CGG	AGA	CAG	CGA	CGA	AGA	GCT	CCT	CAG	180
Ser	Tyr	Gly	Arg	Lys	Lys	Arg	Arg	Gln	Arg	Arg	Arg	Ala	Pro	Gln	
				50					55					60	
GAC	AGT	CAG	ACT	CAT	CAA	GCT	TCT	CTA	TCA	AAG	CAA	CCC	GCC	TCC	225
Asp	Ser	Gln	Thr	His	Gln	Ala	Ser	Leu	Ser	Lys	Gln	Pro	Ala	Ser	
				65					70					75	
CAG	TCC	CGA	GGG	GAC	CCG	ACA	GGC	CCG	ACG	GAA	TCG	AAG	AAG	AAG	270
Gln	Ser	Arg	Gly	Asp	Pro	Thr	Gly	Pro	Thr	Glu	Ser	Lys	Lys	Lys	
				80					85					90	
GTG	GAG	AGA	GAG	ACA	GAG	ACA	GAT	CCG	TTC	GAT	TAG				306
Val	Glu	Arg	Glu	Thr	Glu	Thr	Asp	Pro	Phe	Asp	***				
				95					100						

<210> 7
 <211> 309
 <212> PRT
 <213> bovine leukemia virus

<400> 7

Met	Ala	Ser	Val	Val	Gly	Trp	Gly	Pro	His	Ser	Leu	His	Ala	Cys	1	5	10	15
Pro	Ala	Leu	Val	Leu	Ser	Asn	Asp	Val	Thr	Ile	Asp	Ala	Trp	Cys	20	25	30	
Pro	Leu	Cys	Gly	Pro	His	Glu	Arg	Leu	Gln	Phe	Glu	Arg	Ile	Asp	35	40	45	
Thr	Thr	His	Thr	Cys	Glu	Thr	His	Arg	Ile	Thr	Trp	Thr	Ala	Asp	50	55	60	
Gly	Arg	Pro	Phe	Gly	Leu	Asn	Gly	Ala	Leu	Phe	Pro	Arg	Leu	His	65	70	75	
Val	Ser	Arg	Asp	Pro	Ala	Pro	Arg	Ala	Arg	Arg	Leu	Trp	Ile	Asn	80	85	90	
Cys	Pro	Leu	Pro	Ala	Val	Arg	Ala	Gln	Pro	Gly	Pro	Val	Ser	Leu	95	100	105	
Ser	Pro	Phe	Glu	Arg	Ser	Pro	Phe	Gln	Pro	Tyr	Gln	Cys	Gln	Leu	110	115	120	
Pro	Ser	Ala	Ser	Ser	Asp	Gly	Cys	Pro	Val	Ile	Gly	His	Gly	Leu	125	130	135	
Leu	Pro	Trp	Asn	Asn	Leu	Val	Thr	His	Pro	Cys	Pro	Arg	Lys	Val	140	145	150	
Leu	Ile	Leu	Asn	Gln	Met	Ala	Asn	Phe	Ser	Leu	Leu	Pro	Pro	Phe	155	160	165	
Asn	Thr	Leu	Leu	Val	Asp	Pro	Leu	Arg	Leu	Ser	Val	Phe	Ala	Pro	170	175	180	
Asp	Thr	Arg	Gly	Ala	Ile	Arg	Tyr	Leu	Ser	Thr	Leu	Leu	Thr	Leu	185	190	195	
Cys	Pro	Ala	Thr	Cys	Ile	Leu	Pro	Leu	Gly	Glu	Pro	Phe	Ser	Pro	200	205	210	
Asn	Val	Pro	Ile	Cys	Arg	Phe	Pro	Arg	Asp	Ser	Asn	Glu	Pro	Pro	215	220	225	
Leu	Ser	Glu	Phe	Glu	Leu	Pro	Leu	Ile	Gln	Thr	Pro	Gly	Leu	Ser	230	235	240	
Trp	Ser	Val	Pro	Ala	Ile	Asp	Leu	Phe	Leu	Thr	Gly	Pro	Pro	Ser	245	250	255	

Pro Cys Asp Arg Leu His Val Trp Ser Ser Pro Gln Ala Leu Gln
260 265 270

Arg Phe Leu His Asp Pro Thr Leu Thr Trp Ser Glu Leu Val Ala
275 280 285

Ser Arg Lys Leu Arg Leu Asp Ser Pro Leu Lys Leu Gln Leu Leu
290 295 300

Glu Asn Glu Trp Leu Ser Arg Leu Phe ***
305

<210> 8
 <211> 353
 <212> PRT
 <213> human T-lymphotropic virus 1

<400> 8

Met	Ala	His	Phe	Pro	Gly	Phe	Gly	Gln	Ser	Leu	Leu	Phe	Gly	Tyr	15
1				5					10						
Pro	Val	Tyr	Val	Phe	Gly	Asp	Cys	Val	Gln	Gly	Asp	Trp	Cys	Pro	30
				20					25						
Ile	Ser	Gly	Gly	Leu	Cys	Ser	Ala	Arg	Leu	His	Arg	His	Ala	Leu	45
				35					40						
Leu	Ala	Thr	Cys	Pro	Glu	His	Gln	Ile	Thr	Trp	Asp	Pro	Ile	Asp	60
				50					55						
Gly	Arg	Val	Ile	Gly	Ser	Ala	Leu	Gln	Phe	Leu	Ile	Pro	Arg	Leu	75
				65					70						
Pro	Ser	Phe	Pro	Thr	Gln	Arg	Thr	Ser	Lys	Thr	Leu	Lys	Val	Leu	90
				80					85						
Thr	Pro	Pro	Ile	Thr	His	Thr	Thr	Pro	Asn	Ile	Pro	Pro	Ser	Phe	105
				95					100						
Leu	Gln	Ala	Met	Arg	Lys	Tyr	Ser	Pro	Phe	Arg	Asn	Gly	Tyr	Met	120
				110					115						
Glu	Pro	Thr	Leu	Gly	Gln	His	Leu	Pro	Thr	Leu	Ser	Phe	Pro	Asp	135
				125					130						
Pro	Gly	Leu	Arg	Pro	Gln	Asn	Leu	Tyr	Thr	Leu	Trp	Gly	Gly	Ser	150
				140					145						
Val	Val	Cys	Met	Tyr	Leu	Tyr	Gln	Leu	Ser	Pro	Pro	Ile	Thr	Trp	165
				155					160						
Pro	Leu	Leu	Pro	His	Val	Ile	Phe	Cys	His	Pro	Gly	Gln	Leu	Gly	180
				170					175						
Ala	Phe	Leu	Thr	Asn	Val	Pro	Tyr	Lys	Arg	Ile	Glu	Glu	Leu	Leu	195
				185					190						
Tyr	Lys	Ile	Ser	Leu	Thr	Thr	Gly	Ala	Leu	Ile	Ile	Leu	Pro	Glu	210
				200					205						
Asp	Cys	Leu	Pro	Thr	Thr	Leu	Phe	Gln	Pro	Val	Arg	Ala	Pro	Val	225
				215					220						
Thr	Leu	Thr	Ala	Trp	Gln	Asn	Gly	Leu	Leu	Pro	Phe	His	Ser	Thr	240
				230					235						
Leu	Thr	Thr	Pro	Gly	Leu	Ile	Trp	Thr	Phe	Thr	Asp	Gly	Thr	Pro	

	245		250		255
Met Ile Ser Gly	Pro Cys Pro Lys Asp	Gly Gln Pro Ser Leu Val			
	260		265		270
Leu Gln Ser Ser	Ser Phe Ile Phe His	Lys Phe Gln Thr Lys Ala			
	275		280		285
Tyr His Pro Ser	Phe Leu Leu Ser His	Gly Leu Ile Gln Tyr Ser			
	290		295		300
Ser Phe His Asn	Leu His Leu Leu Phe	Glu Glu Tyr Thr Asn Ile			
	305		310		315
Pro Ile Ser Leu	Leu Phe Asn Lys Lys	Glu Ala Asp Asp Asn Asp			
	320		325		330
His Glu Pro Gln	Ile Ser Pro Gly Gly	Leu Glu Pro Pro Ser Glu			
	335		340		345
Lys His Phe Arg	Glu Thr Glu Val ***				
	350				

<210> 9
<211> 101
<212> PRT
<213> human immunodeficiency virus 1

<400> 9

Met	Glu	Pro	Val	Asp	Pro	Asn	Leu	Glu	Pro	Trp	Lys	His	Pro	Gly	
1				5					10					15	
Ser	Gln	Pro	Arg	Thr	Ala	Cys	Asn	Asn	Cys	Tyr	Cys	Lys	Lys	Cys	
				20					25					30	
Cys	Phe	His	Cys	Tyr	Ala	Cys	Phe	Thr	Arg	Lys	Gly	Leu	Gly	Ile	
				35					40					45	
Ser	Tyr	Gly	Arg	Lys	Lys	Arg	Arg	Gln	Arg	Arg	Arg	Ala	Pro	Gln	
				50					55					60	
Asp	Ser	Gln	Thr	His	Gln	Ala	Ser	Leu	Ser	Lys	Gln	Pro	Ala	Ser	
				65					70					75	
Gln	Ser	Arg	Gly	Asp	Pro	Thr	Gly	Pro	Thr	Glu	Ser	Lys	Lys	Lys	
				80					85					90	
Val	Glu	Arg	Glu	Thr	Glu	Thr	Asp	Pro	Phe	Asp	***				
				95					100						

<210> 10
<211> 7685
<212> DNA
<213> Artificial Sequence

<220>
<221> CDS
<222> 1753..2148
<223> Blasticidin Resistance

<220>
<221> CDS
<222> 3115..4041
<223> BLV Tax

<220>
<221> CDS
<222> 6616..7476
<223> Ampicillin Resistance

<220>
<221> LTR
<222> 149..737
<223> 5' MoMuSVLTR

<220>
<221> LTR
<222> 4720..5313
<223> 3' MoMuLVLTR

<220>
<221> misc_recomb
<222> 3078..3102
<223> attB1

<220>
<221> misc_recomb
<222> 4046..4070
<223> attB2

<220>
<221> misc_signal
<222> 4082..4674
<223> WPRE; woodchuck hepatitis virus post-transcriptional regulatory element

<220>
<221> promoter
<222> 2257..3074
<223> CMV IE promoter

<400> 10

GAATTAATTC ATACCAGATC ACCGAAAAC TGCCTCCAAA TGTGTCCCC TCACACTCCC 60

AAATTCGCGG GCTTCTGCCT CTTAGACCAC TCTACCCTAT TCCCCACACT CACCGGAGCC 120

AAAGCCGCGG CCCTTCCGTT TCTTTGCTTT TGAAAGACCC CACCCGTAGG TGGCAAGCTA 180
GCTTAAGTAA CGCCACTTTG CAAGGCATGG AAAAATACAT AACTGAGAAT AGAAAAGTTC 240
AGATCAAGGT CAGGAACAAA GAAACAGCTG AATACCAAAC AGGATATCTG TGGTAAGCGG 300
TTCCTGCCCC GGCTCAGGGC CAAGAACAGA TGAGACAGCT GAGTGATGGG CCAAACAGGA 360
TATCTGTGGT AAGCAGTTCC TGCCCCGGCT CGGGGCCAAG AACAGATGGT CCCCAGATGC 420
GGTCCAGCCC TCAGCAGTTT CTAGTGAATC ATCAGATGTT TCCAGGGTGC CCCAAGGACC 480
TGAAAATGAC CCTGTACCTT ATTTGAACTA ACCAATCAGT TCGCTTCTCG CTTCTGTTCG 540
CGCGCTTCCG CTCTCCGAGC TCAATAAAAG AGCCCACAAC CCCTCACTCG GCGCGCCAGT 600
CTTCCGATAG ACTGCGTCGC CCGGGTACCC GTATTCCCAA TAAAGCCTCT TGCTGTTTGC 660
ATCCGAATCG TGGTCTCGCT GTTCCTTGGG AGGGTCTCCT CTGAGTGATT GACTACCCAC 720
GACGGGGGTC TTTCATTTGG GGGCTCGTCC GGGATTTGGA GACCCCTGCC CAGGGACCAC 780
CGACCCACCA CCGGGAGGTA AGCTGGCCAG CAACTTATCT GTGTCTGTCC GATTGTCTAG 840
TGCTCTATGTT TGATGTTATG CGCCTGCGTC TGTACTAGTT AGCTAACTAG CTCTGTATCT 900
GGCGGACCCG TGGTGGAAC TACGAGTTCT GAACACCCGG CCGCAACCCT GGGAGACGTC 960
CCAGGGACTT TGGGGGCCGT TTTTGTGGCC CGACCTGAGG AAGGGAGTCG ATGTGGAATC 1020
CGACCCCGTC AGGATATGTG GTTCTGGTAG GAGACGAGAA CCTAAAACAG TTCCCGCCTC 1080
CGTCTGAATT TTTGCTTTTCG GTTTGGAACC GAAGCCGCGC GTCTTGTCTG CTGCAGCGCT 1140
GCAGCATCGT TCTGTGTTGT CTCTGTCTGA CTGTGTTTCT GTATTTGTCT GAAAATTAGG 1200
GCCAGACTGT TACCACTCCC TTAAGTTTGA CCTTAGGTCA CTGGAAAGAT GTCGAGCGGA 1260
TCGCTCACAA CCAGTCGGTA GATGTCAAGA AGAGACGTTG GGTACCTTC TGCTCTGCAG 1320
AATGGCCAAC CTTTAACGTC GGATGGCCGC GAGACGGCAC CTTTAACCGA GACCTCATCA 1380
CCCAGGTAA GATCAAGGTC TTTTCACCTG GCCCGCATGG ACACCCAGAC CAGGTCCCCT 1440
ACATCGTGAC CTGGGAAGCC TTGGCTTTTG ACCCCCCTCC CTGGGTCAAG CCCTTTGTAC 1500
ACCCTAAGCC TCCGCCTCCT CTTCTCCAT CCGCCCCGTC TCTCCCCCTT GAACCTCCTC 1560
GTTCGACCCC GCCTCGATCC TCCCTTTATC CAGCCCTCAC TCCTTCTCTA GGCGCCGGAA 1620
TTCCGATCTG ATCAAGAGAC AGGATGAGGG AGCTTGTATA TCCATTTTCG GATCTGATCA 1680
GCACGTGTTG ACAATTAATC ATCGGCATAG TATATCGGCA TAGTATAATA CGACAAGGTG 1740
AGGAACTAAA CCATGGCCAA GCCTTTGTCT CAAGAAGAAT CCACCCTCAT TGAAAGAGCA 1800
ACGGCTACAA TCAACAGCAT CCCCATCTCT GAAGACTACA GCGTCGCCAG CGCAGCTCTC 1860

TCTAGCGACG GCCGCATCTT CACTGGTGTC AATGTATATC ATTTTACTGG GGGACCTTGT 1920
GCAGAACTCG TGGTGCTGGG CACTGCTGCT GCTGCGGCAG CTGGCAACCT GACTTGTATC 1980
GTCGCGATCG GAAATGAGAA CAGGGGCATC TTGAGCCCCCT GCGGACGGTG TCGACAGGTG 2040
CTTCTCGATC TGCATCCTGG GATCAAAGCG ATAGTGAAGG ACAGTGATGG ACAGCCGACG 2100
GCAGTTGGGA TTCGTGAATT GCTGCCCTCT GGTTATGTGT GGGAGGGCTA AGCACTTCGT 2160
GGCCGAGGAG CAGGACTGAC ACGTGCTACG AGATTTTCGAT TCCACCGCCG CCTTCTATGA 2220
AAGGTTGGGC TTCGGAATCG TTTTCCGGGA CGCCGATCCG GCCATTAGCC ATATTATTCA 2280
TTGGTTATAT AGCATAAATC AATATTGGCT ATTGGCCATT GCATACGTTG TATCCATATC 2340
ATAATATGTA CATTTATATT GGCTCATGTC CAACATTACC GCCATGTTGA CATTGATTAT 2400
TGACTAGTTA TTAATAGTAA TCAATTACGG GGTCATTAGT TCATAGCCCA TATATGGAGT 2460
TCCGCGTTAC ATAACCTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCGGCC 2520
CATTGACGTC AATAATGACG TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC 2580
GTCAATGGGT GGAGTATTTA CGGTAAACTG CCCACTTGGC AGTACATCAA GTGTATCATA 2640
TGCCAAGTAC GCCCCCTATT GACGTCAATG ACGGTAAAATG GCCCGCCTGG CATTATGCCC 2700
AGTACATGAC CTTATGGGAC TTTCCTACTT GGCAGTACAT CTACGTATTA GTCATCGCTA 2760
TTACCATGGT GATGCGGTTT TGGCAGTACA TCAATGGGCG TGGATAGCGG TTTGACTCAC 2820
GGGGATTTCC AAGTCTCCAC CCCATTGACG TCAATGGGAG TTTGTTTTGG CACCAAAATC 2880
AACGGGACTT TCCAAAATGT CGTAACAACCT CCGCCCCATT GACGCAAATG GGCGGTAGGC 2940
ATGTACGGTG GGAGGTCTAT ATAAGCAGAG CTCGTTTAGT GAACCGTCAG ATCGCCTGGA 3000
GACGCCATCC ACGCTGTTTT GACCTCCATA GAAGACACCG GGACCGATCC AGCCTCCGCG 3060
GCCCCAAGCT TGTTATCACA AGTTTGTACA AAAAAGCAGG CTCCCGCCGC CACC ATG 3117
Met
1
GCA AGT GTT GTT GGT TGG GGG CCC CAC TCT CTA CAT GCC TGC CCG GCC 3165
Ala Ser Val Val Gly Trp Gly Pro His Ser Leu His Ala Cys Pro Ala
5 10 15
CTG GTT TTG TCC AAT GAT GTC ACC ATC GAT GCC TGG TGC CCC CTC TGC 3213
Leu Val Leu Ser Asn Asp Val Thr Ile Asp Ala Trp Cys Pro Leu Cys
20 25 30
GGG CCC CAT GAG CGA CTC CAA TTC GAA AGG ATC GAC ACC ACG CTC ACC 3261
Gly Pro His Glu Arg Leu Gln Phe Glu Arg Ile Asp Thr Thr Leu Thr
35 40 45

TGC GAG ACC CAC CGT ATC AAC TGG ACC GCC GAT GGA CGA CCT TGC GGC	3309
Cys Glu Thr His Arg Ile Asn Trp Thr Ala Asp Gly Arg Pro Cys Gly	
50 55 60 65	
CTC AAT GGA ACG TTG TTC CCT CGA CTG CAT GTC TCC GAG ACC CGC CCC	3357
Leu Asn Gly Thr Leu Phe Pro Arg Leu His Val Ser Glu Thr Arg Pro	
70 75 80	
CAA GGG CCC CGA CGA CTC TGG ATC AAC TGC CCC CTT CCG GCC GTT CGC	3405
Gln Gly Pro Arg Arg Leu Trp Ile Asn Cys Pro Leu Pro Ala Val Arg	
85 90 95	
GCT CAG CCC GGC CCG GTT TCA CTT TCC CCC TTC GAG CGG TCC CCC TTC	3453
Ala Gln Pro Gly Pro Val Ser Leu Ser Pro Phe Glu Arg Ser Pro Phe	
100 105 110	
CAG CCC TAC CAA TGC CAA TTG CCC TCG GCC TCT AGC GAC GGT TGC CCC	3501
Gln Pro Tyr Gln Cys Gln Leu Pro Ser Ala Ser Ser Asp Gly Cys Pro	
115 120 125	
ATT ATC GGG CAC GGC CTT CTT CCC TGG AAC AAC TTA GTA ACG CAT CCT	3549
Ile Ile Gly His Gly Leu Leu Pro Trp Asn Asn Leu Val Thr His Pro	
130 135 140 145	
GTC CTC AGA AAA GTC CTT ATA TTA AAT CAA ATG GCC AAT TTT TCC TTA	3597
Val Leu Arg Lys Val Leu Ile Leu Asn Gln Met Ala Asn Phe Ser Leu	
150 155 160	
CTC CCC TCC TTC GAT ACC CTC CTT GTG GAC CCC CTC CGG CTG TCC GTC	3645
Leu Pro Ser Phe Asp Thr Leu Leu Val Asp Pro Leu Arg Leu Ser Val	
165 170 175	
TTT GCC CCA GAC ACC AGG GGA GCC ATA CGT TAT CTC TCC ACC CTT TTG	3693
Phe Ala Pro Asp Thr Arg Gly Ala Ile Arg Tyr Leu Ser Thr Leu Leu	
180 185 190	
ACG CTA TGC CCG GCT ACT TGT ATT CTA CCC CTA GGC GAG CCC TTC TCT	3741
Thr Leu Cys Pro Ala Thr Cys Ile Leu Pro Leu Gly Glu Pro Phe Ser	
195 200 205	
CCT AAT GTC CCC ATA TGC CGC TTT CCC CGG GAC TCC AAT GAA CCC CCC	3789
Pro Asn Val Pro Ile Cys Arg Phe Pro Arg Asp Ser Asn Glu Pro Pro	
210 215 220 225	
CTT TCA GAA TTC GAG CTG CCC CTT ATC CAA ACG CCC GGC CTG TCT TGG	3837
Leu Ser Glu Phe Glu Leu Pro Leu Ile Gln Thr Pro Gly Leu Ser Trp	
230 235 240	
TCT GTC CCC GCG ATC GAC CTA TTC CTA ACC GGT CCC CCT TCC CCA TGC	3885
Ser Val Pro Ala Ile Asp Leu Phe Leu Thr Gly Pro Pro Ser Pro Cys	
245 250 255	
GAC CGG TTA CAC GTA TGG TCC AGT CCT CAG GCC TTA CAG CGC TTC CTT	3933
Asp Arg Leu His Val Trp Ser Ser Pro Gln Ala Leu Gln Arg Phe Leu	
260 265 270	
CAT GAC CCT ACG CTA ACC TGG TCC GAA TTA GTT GCT AGC AGA AAA ATA	3981

His Asp Pro Thr Leu Thr Trp Ser Glu Leu Val Ala Ser Arg Lys Ile
 275 280 285

AGA CTT GAT TCC CCC TTA AAA TTA CAA CTG CTA GAA AAT GAA TGG CTC 4029
 Arg Leu Asp Ser Pro Leu Lys Leu Gln Leu Leu Glu Asn Glu Trp Leu
 290 295 300 305

TCC CGC CTT TTT TGA GACCCA GCTTTCTTGT ACAAAGTGGT GATAACATCG 4080
 Ser Arg Leu Phe ***

ATAATCAACC TCTGGATTAC AAAATTTGTG AAAGATTGAC TGGTATTCTT AACTATGTTG 4140
 CTCCTTTTAC GCTATGTGGA TACGCTGCTT TAATGCCTTT GTATCATGCT ATTGCTTCCC 4200
 GTATGGCTTT CATTTTCTCC TCCTTGTATA AATCCTGGTT GCTGTCTCTT TATGAGGAGT 4260
 TGTGGCCCGT TGTCAGGCAA CGTGGCGTGG TGTGCACTGT GTTTGCTGAC GCAACCCCCA 4320
 CTGGTTGGGG CATTGCCACC ACCTGTCAGC TCCTTTCCGG GACTTTCGCT TTCCCCCTCC 4380
 CTATTGCCAC GCGGAACTC ATCGCCGCTT GCCTTGCCCG CTGCTGGACA GGGGCTCGGC 4440
 TGTGGGCAC TGACAATTCC GTGGTGTGTG CGGGGAAATC ATCGTCCTTT CCTTGGCTGC 4500
 TCGCCTGTGT TGCCACCTGG ATTCTGCGCG GGACGTCTTT CTGCTACGTC CCTTCGGCCC 4560
 TCAATCCAGC GGACCTTCCT TCCCGCGGCC TGCTGCCGGC TCTGCGGCCT CTTCCGCGTC 4620
 TTCGCCCTTCG CCCTCAGACG AGTCGGATCT CCCTTTGGGC CGCCTCCCCG CCTGATCGAT 4680
 AAAATAAAAG ATTTTATTTA GTCTCCAGAA AAAGGGGGGA ATGAAAGACC CCACCTGTAG 4740
 GTTTGGCAAG CTAGCTTAAG TAACGCCATT TTGCAAGGCA TGGAAAAATA CATAACTGAG 4800
 AATAGAGAAG TTCAGATCAA GGTGAGGAAC AGATGGAACA GCTGAATATG GGCCAAACAG 4860
 GATATCTGTG GTAAGCAGTT CCTGCCCCGG CTCAGGGCCA AGAACAGATG GAACAGCTGA 4920
 ATATGGGCCA AACAGGATAT CTGTGGTAAG CAGTTCCTGC CCCGGCTCAG GGCCAAGAAC 4980
 AGATGGTCCC CAGATGCGGT CCAGCCCTCA GCAGTTTCTA GAGAACCATC AGATGTTTCC 5040
 AGGGTGCCCC AAGGACCTGA AATGACCCTG TGCCTTATTT GAACTAACCA ATCAGTTTCG 5100
 TTCTCGCTTC TGTTGCGCGC CTTCTGCTCC CCGAGCTCAA TAAAAGAGCC CACAACCCCT 5160
 CACTCGGGGC GCCAGTCCTC CGATTGACTG AGTCGCCCCG GTACCCGTGT ATCCAATAAA 5220
 CCCTCTTGCA GTTGCATCCG ACTTGTGGTC TCGCTGTTCC TTGGGAGGGT CTCCTCTGAG 5280
 TGATTGACTA CCCGTCAGCG GGGGTCTTTC ATTTTTCAT TGGGGGCTCG TCCGGGATCG 5340
 GGAGACCCCT GCCCAGGGAC CACCGACCCA CCACGGGAG GTAAGCTGGC TGCCTCGCGC 5400
 GTTTCGGTGA TGACGGTGAA AACCTCTGAC ACATGCAGCT CCCGGAGACG GTCACAGCTT 5460
 GTCTGTAAGC GGATGCCGGG AGCAGACAAG CCCGTCAGGG CGCGTCAGCG GGTGTTGGCG 5520

GGTGTCTGGGG CGCAGCCATG ACCCAGTCAC GTAGCGATAG CGGAGTGTAT ACTGGCTTAA 5580
CTATGCGGCA TCAGAGCAGA TTGTACTGAG AGTGCACCAT ATGCGGTGTG AAATACCGCA 5640
CAGATGCGTA AGGAGAAAAT ACCGCATCAG GCGCTCTTCC GCTTCCTCGC TCACTGACTC 5700
GCTGCGCTCG GTCGTTCGGC TCGGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG 5760
GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA 5820
GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCCTTTTTT CATAGGCTCC GCCCCCTGA 5880
CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG 5940
ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT 6000
TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGAAGCGTG GCGCTTTCTC ATAGCTCAG 6060
CTGTAGGTAT CTCAGTTCCG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC 6120
CCCCGTTTCA CCCGACCGCT GCGCCTTATC CGGTAAGTAT CGTCTTGAGT CCAACCCGGT 6180
AAGACACGAC TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA 6240
TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGGAC 6300
AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC 6360
TTGATCCGGC AAACAAACCA CCGCTGGTAG CCGTGGTTTT TTTGTTTGCA AGCAGCAGAT 6420
TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC 6480
TCAGTGGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT 6540
CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA 6600
AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT 6660
ATTTCTGTTCA TCCATAGTTG CCTGACTCCC CGTCGTGTAG ATAACCTACGA TACGGGAGGG 6720
CTTACCATCT GGCCCCAGTG CTGCAATGAT ACCGCGAGAC CCACGCTCAC CGGCTCCAGA 6780
TTTATCAGCA ATAAACCAGC CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC CTGCAACTTT 6840
ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGAAGCT AGAGTAAGTA GTTCGCCAGT 6900
TAATAGTTTG CGCAACGTTG TTGCCATTGC TGCAGGCATC GTGGTGTAC GCTCGTCGTT 6960
TGGTATGGCT TCATTCAGCT CCGGTTCCCA ACGATCAAGG CGAGTTACAT GATCCCCCAT 7020
GTTGTGCAAA AAAGCGGTTA GCTCCTTCGG TCCTCCGATC GTTGTCAGAA GTAAGTTGGC 7080
CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC 7140
CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG AATAGTGTAT 7200

GCGGCGACCG AGTTGCTCTT GCCCGGCGTC AACACGGGAT AATACCGCGC CACATAGCAG 7260
AACTTTAAAA GTGCTCATCA TTGGAAAACG TTCTTCGGGG CGAAACTCT CAAGGATCTT 7320
ACCGCTGTTG AGATCCAGTT CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC 7380
TTTTACTTTC ACCAGCGTTT CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA 7440
GGGAATAAGG GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTT AATATTATTG 7500
AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA 7560
TAAACAAATA GGGGTTCGCG GCACATTTCC CCGAAAAGTG CCACCTGACG TCTAAGAAAC 7620
CATTATTATC ATGACATTAA CCTATAAAAA TAGGCGTATC ACGAGGCCCT TTCGTCTTCA 7680
AGAAT 7685

<210> 11
 <211> 7430
 <212> DNA
 <213> artificial sequence

<220>
 <221> CDS
 <222> 3120..3590
 <223> trans-dominant BLV Rex (M4)

<220>
 <221> CDS
 <222> 1512..2306
 <223> neomycin resistance

<220>
 <221> CDS
 <222> 6217..7077
 <223> ampicillin resistance

<220>
 <221> LTR
 <222> 1..589
 <223> 5' MoMuSVLTR

<220>
 <221> LTR
 <222> 4328..4921
 <223> 3' MoMuLVLTR

<220>
 <221> misc_feature
 <222> 3023..3047
 <223> attB1

<220>
 <221> misc_feature
 <222> 3653..4282
 <223> attB2

<220>
 <221> misc_signal
 <222> 3690..4282
 <223> WPRE; woodchuck hepatitis virus post-transcriptional regulatory element

<400> 11

TTTGAAAGAC CCCACCCGTA GGTGGCAAGC TAGCTTAAGT AACGCCACTT TGCAAGGCAT 60
 GGAAAAATAC ATAAGTGAGA ATAGAAAAGT TCAGATCAAG GTCAGGAACA AAGAAACAGC 120
 TGAATACCAA ACAGGATATC TGTGGTAAGC GGTTCCTGCC CCGGCTCAGG GCCAAGAACA 180
 GATGAGACAG CTGAGTGATG GGCCAAACAG GATATCTGTG GTAAGCAGTT CCTGCCCCGG 240
 CTCGGGGCCA AGAACAGATG GTCCCCAGAT GCGGTCCAGC CCTCAGCAGT TTCTAGTGAA 300

TCATCAGATG TTTCCAGGGT GCCCCAAGGA CCTGAAAATG ACCCTGTACC TTATTTGAAC 360
TAACCAATCA GTTCGCTTCT CGCTTCTGTT CGCGCGCTTC CGCTCTCCGA GCTCAATAAA 420
AGAGCCCACA ACCCCTCACT CGGCGCGCCA GTCTTCCGAT AGACTGCGTC GCCCGGGTAC 480
CCGTATTCCC AATAAAGCCT CTTGCTGTTT GCATCCGAAT CGTGGTCTCG CTGTTCTTG 540
GGAGGGTCTC CTCTGAGTGA TTGACTACCC ACGACGGGGG TCTTTCATTT GGGGGCTCGT 600
CCGGGATTTG GAGACCCCTG CCCAGGGACC ACCGACCCAC CACCGGGAGG TAAGCTGGCC 660
AGCAACTTAT CTGTGTCTGT CCGATTGTCT AGTGTCTATG TTTGATGTTA TGCGCCTGCG 720
TCTGTACTAG TTAGCTAACT AGCTCTGTAT CTGGCGGACC CGTGGTGGAA CTGACGAGTT 780
CTGAACACCC GGCCGCAACC CTGGGAGACG TCCCAGGGAC TTTGGGGGCC GTTTTTGTGG 840
CCCGACCTGA GGAAGGGAGT CGATGTGGAA TCCGACCCCG TCAGGATATG TGGTTCTGGT 900
AGGAGACGAG AACCTAAAC AGTCCCCGCC TCCGTCTGAA TTTTGTCTTT CGGTTTGGAA 960
CCGAAGCCGC GCGTCTTGTC TGCTGCAGCG CTGCAGCATC GTTCTGTGTT GTCTCTGTCT 1020
GACTGTGTTT CTGTATTTGT CTGAAAATTA GGGCCAGACT GTTACCACTC CCTTAAGTTT 1080
GACCTTAGGT CACTGGAAAG ATGTCGAGCG GATCGCTCAC AACCAGTCGG TAGATGTCAA 1140
GAAGAGACGT TGGGTTACCT TCTGCTCTGC AGAATGGCCA ACCTTTAACG TCGGATGGCC 1200
GCGAGACGGC ACCTTTAACC GAGACCTCAT CACCCAGGTT AAGATCAAGG TCTTTTCACC 1260
TGGCCCGCAT GGACACCCAG ACCAGGTCCC CTACATCGTG ACCTGGGAAG CCTTGGCTTT 1320
TGACCCCCCT CCCTGGGTCA AGCCCTTTGT ACACCCTAAG CCTCCGCCTC CTCTTCCTCC 1380
ATCCGCCCCG TCTCTCCCCC TTGAACCTCC TCGTTCGACC CCGCTCGAT CCTCCCTTTA 1440
TCCAGCCCTC ACTCCTTCTC TAGGCGCCGG AATTCCGATC TGATCAAGAG ACAGGATGAG 1500
GATCGTTTCG CATGATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC GCTTGGGTGG 1560
AGAGGCTATT CGGCTATGAC TGGGCACAAC AGACAATCGG CTGCTCTGAT GCCGCCGTGT 1620
TCCGGCTGTC AGCGCAGGGG CGCCCGGTTT TTTTGTCAA GACCGACCTG TCCGGTGCCC 1680
TGAATGAACT GCAGGACGAG GCAGCGCGGC TATCGTGGCT GGCCACGACG GGC GTTCCTT 1740
GCGCAGCTGT GCTCGACGTT GTCAGTGAAG CGGGAAGGGA CTGGCTGCTA TTGGGCGAAG 1800
TGCCGGGGCA GGATCTCCTG TCATCTCACC TTGCTCCTGC CGAGAAAGTA TCCATCATGG 1860
CTGATGCAAT GCGGCGGCTG CACGCTTG ATCCGGCTAC CTGCCCATTC GACCACCAAG 1920
CGAAACATCG CATCGAGCGA GCACGTACTC GGATGGAAGC CGGTCTTGTC GATCAGGATG 1980

ATCTGGACGA AGAGCATCAG GGGCTCGCGC CAGCCGAACT GTTCGCCAGG CTCAAGGCGC 2040
 GCATGCCCCGA CGGCGAGGAT CTCGTCGTGA CCCATGGCGA TGCCTGCTTG CCGAATATCA 2100
 TGGTGGAAAA TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT GTGGCGGACC 2160
 GCTATCAGGA CATAGCGTTG GCTACCCGTG ATATTGCTGA AGAGCTTGGC GGCGAATGGG 2220
 CTGACCGCTT CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC ATCGCCTTCT 2280
 ATCGCCTTCT TGACGAGTTC TTCTGAGCGG GACTCTGGGG TTCGAAATGA CCGACCAAGC 2340
 GACGCCCAAC CTGCCATCAC GAGATTTCTGA TTCCACCGCC GCCTTCTATG AAAGGTGGG 2400
 CTTCGGAATC GTTTTCCGGG ACGCCGGCTG GATGATCCTC CAGCGCGGGG ATCTCATGCT 2460
 GGAGTTCTTC GCCCACCCCG GGCTCGATCC CCTCGCGAGT TGGTTCAGCT GCTGCCTGAG 2520
 GCTGGACGAC CTCGCGGAGT TCTACCGGCA GTGCAAATCC GTCGGCATCC AGGAAACCAG 2580
 CAGCGGCTAT CCGCGCATCC ATGCCCCCGA ACTGCAGGAG TGGGGAGGCA CGATGGCCGC 2640
 TTTGGTCGAG GCGGATCCTA GCAGAAAAAT AAGACTTGAT TCCCCCTTAA AATTACAAC 2700
 GCTAGAAAAT GAATGGCTCT CCCGCCTTTT TTGAGGGGGA ATCATTTGTA TGAAAGATCA 2760
 TGCCGACCTA GGCGCCGCCA CCGCCCCGTA AACCAGACAG AGACGTCAGC TGCCAGAAAA 2820
 GCTGGTGACG GCAGCTGGTG GCTAGAATCC CCGTACCTCC CCAACTTCCC CTTTCCCGAA 2880
 AAATCCACAC CCTGAGCTGC TGACCTCACC TGCTGATAAA TTAATAAAAT GCCGGCCCTG 2940
 TCGAGTTAGC GGCACCAGAA GCGTTCCTCT CCTGAGACCC TCGTGCTCAG CTCTCGGTCC 3000
 TGCCTCGAGA AGCTTGTTAT CAACAAGTTT GTACAAAAAA GCAGGCTTCG AAGGAGATAG 3060
 AACCAATTCT CTAAGGAAAT ACTTAACGTC GACTGGATCC GGTACCGAAT TCGATCCAC 3119
 ATG CCT AAA AAA CGA CGG TCC CGA AGA CGC CCA CAA CCG ATC ATC 3164
 Met Pro Lys Lys Arg Arg Ser Arg Arg Arg Pro Gln Pro Ile Ile
 1 5 10 15
 AGA TGG CAA GTG TTG TTG GTT GGG GGC CCC ACT CTC TAC ATG CCT 3209
 Arg Trp Gln Val Leu Leu Val Gly Gly Pro Thr Leu Tyr Met Pro
 20 25 30
 GCC CGG CCC TGG TTT TGT CCA ATG ATG TCA CCA TCG ATG CCT GGT 3254
 Ala Arg Pro Trp Phe Cys Pro Met Met Ser Pro Ser Met Pro Gly
 35 40 45
 GCC CCC TCT GCG GGC CCC ATG AGC GAC TCC AAT TCG AAA GGA TCG 3299
 Ala Pro Ser Ala Gly Pro Met Ser Asp Ser Asn Ser Lys Gly Ser
 50 55 60
 ACA CCA CGC TCA CCT GCG AGA CCC ACC GTA TCA ACT GGA CCG CCG 3344
 Thr Pro Arg Ser Pro Ala Arg Pro Thr Val Ser Thr Gly Pro Pro
 65 70 75

ATG GAC GAC CTT GCG GCC TCA ATG GAA CGT TGT TCC CTC GAC TGC	3389
Met Asp Asp Leu Ala Ala Ser Met Glu Arg Cys Ser Leu Asp Cys	
80 85 90	
ATG TCT CCG AGA CCC GCC CCC AAG GGC CCC GAC GAC TCT GGA TCA	3434
Met Ser Pro Arg Pro Ala Pro Lys Gly Pro Asp Asp Ser Gly Ser	
95 100 105	
ACT GCC CCC TTC CGG CCG TTC GCG CTC AGC CCG GCC CGG TTA GAT	3479
Thr Ala Pro Phe Arg Pro Phe Ala Leu Ser Pro Ala Arg Leu Asp	
110 115 120	
CTT CCC CCT TCG AGC GGT CCC CCT TCC AGC CCT ACC AAT GCC AAT	3524
Leu Pro Pro Ser Ser Gly Pro Pro Ser Ser Pro Thr Asn Ala Asn	
125 130 135	
TGC CCT CGG CCT CTA GCG ACG GTT GCC CCA TTA TCG GGC ACG GCC	3569
Cys Pro Arg Pro Leu Ala Thr Val Ala Pro Leu Ser Gly Thr Ala	
140 145 150	
TTC TTC CCT GGA ACA ACT TAG TAACGCATCC	3600
Phe Phe Pro Gly Thr Thr ***	
155	
TGTCCTCAGA AAAGTCCTTA TATTAAATCA AATGGGACCT CGAGATATCT AGACCCAGCT	3660
TTCTTGTACA AAGTGGTTGA TAACATCGAT AATCAACCTC TGGATTACAA AATTTGTGAA	3720
AGATTGACTG GTATTCTTAA CTATGTTGCT CCTTTTACGC TATGTGGATA CGCTGCTTTA	3780
ATGCCTTTGT ATCATGCTAT TGCTTCCCGT ATGGCTTTCA TTTTCTCCTC CTTGTATAAA	3840
TCCTGGTTGC TGTCTCTTTA TGAGGAGTTG TGGCCCGTTG TCAGGCAACG TGGCGTGGTG	3900
TGCACTGTGT TTGCTGACGC AACCCCCACT GGTGGGGCA TTGCCACCAC CTGTGACGTC	3960
CTTTCCGGGA CTTTCGCTTT CCCCCTCCCT ATTGCCACGG CGGAACTCAT CGCCGCCTGC	4020
CTTGCCCGCT GCTGGACAGG GGCTCGGCTG TTGGGCACTG ACAATTCCGT GGTGTTGTCTG	4080
GGGAAATCAT CGTCCTTTCC TTGGCTGCTC GCCTGTGTTG CCACCTGGAT TCTGCGCGGG	4140
ACGTCCTTCT GCTACGTCCC TTCGGCCCTC AATCCAGCGG ACCTTCCTTC CCGCGGCCTG	4200
CTGCCGGCTC TCGGCCTCT TCCGCGTCTT CGCCTTCGCC CTCAGACGAG TCGGATCTCC	4260
CTTTGGGCCG CCTCCCCGCC TGATCGATAA AATAAAAGAT TTTATTTAGT CTCCAGAAAA	4320
AGGGGGGAAT GAAAGACCCC ACCTGTAGGT TTGGCAAGCT AGCTTAAGTA ACGCCATTTT	4380
GCAAGGCATG GAAAAATACA TAACTGAGAA TAGAGAAGTT CAGATCAAGG TCAGGAACAG	4440
ATGGAACAGC TGAATATGGG CCAAACAGGA TATCTGTGGT AAGCAGTTCC TGCCCCGGCT	4500
CAGGGCCAAG AACAGATGGA ACAGCTGAAT ATGGGCCAAA CAGGATATCT GTGGTAAGCA	4560

GTTCCTGCCC CGGCTCAGGG CCAAGAACAG ATGGTCCCCA GATGCGGTCC AGCCCTCAGC 4620
 AGTTTCTAGA GAACCATCAG ATGTTTCCAG GGTGCCCCAA GGACCTGAAA TGACCCTGTG 4680
 CCTTATTTGA ACTAACCAAT CAGTTCGCTT CTCGCTTCTG TTCGCGCGCT TCTGCTCCCC 4740
 GAGCTCAATA AAAGAGCCCA CAACCCCTCA CTCGGGGCGC CAGTCCTCCG ATTGACTGAG 4800
 TCGCCCGGGT ACCCGTGTAT CCAATAAACC CTCTTGCACT TGCATCCGAC TTGTGGTCTC 4860
 GCTGTTCCCTT GGGAGGGTCTT CCTCTGAGTG ATTGACTACC CGTCAGCGGG GGTCTTTCAT 4920
 TTGGGGGCTC GTCCGGGATC GGGAGACCCC TGCCCAGGGA CCACCGACCC ACCACCGGGA 4980
 GGTAAGCTGG CTGCCTCGCG CGTTTCGGTG ATGACGGTGA AAACCTCTGA CACATGCAGC 5040
 TCCCGGAGAC GGTACAGCT TGTCTGTAAG CGGATGCCGG GAGCAGACAA GCCCGTCAGG 5100
 GCGCGTCAGC GGGTGTGGC GGGTGTGGG GCGCAGCCAT GACCCAGTCA CGTAGCGATA 5160
 GCGGAGTGTA TACTGGCTTA ACTATGCGGC ATCAGAGCAG ATTGTACTGA GAGTGCACCA 5220
 TATGCGGTGT GAAATACCGC ACAGATGCGT AAGGAGAAAA TACCGCATCA GCGCTCTTC 5280
 CGCTTCCTCG CTCACTGACT CGCTGCGCTC GGTTCGTCGG CTGCGGCGAG CGGTATCAGC 5340
 TCACTCAAAG GCGGTAATAC GGTATCCAC AGAATCAGGG GATAACGCAG GAAAGAACAT 5400
 GTGAGCAAAA GGCCAGCAA AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC TGGCGTTTTT 5460
 CCATAGGCTC CGCCCCCTG ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG 5520
 AAACCCGACA GGA CTATAAA GATACCAGGC GTTTCCCCCT GGAAGCTCCC TCGTGCGCTC 5580
 TCCTGTTCG ACCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAGCGT 5640
 GCGCTTTCT CATAGCTCAC GCTGTAGGTA TCTCAGTTCG GTGTAGGTCG TTCGCTCCAA 5700
 GCTGGGCTGT GTGCACGAAC CCCCCGTTCA GCCCGACCGC TGC GCCTTAT CCGGTA ACTA 5760
 TCGTCTTGAG TCCAACCCGG TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACTGGTAA 5820
 CAGGATTAGC AGAGCGAGGT ATGTAGGCGG TGCTACAGAG TTCTTGAAGT GGTGGCCTAA 5880
 CTACGGCTAC ACTAGAAGGA CAGTATTTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT 5940
 CGGAAAAAGA GTTGGTAGCT CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTTT 6000
 TTTTGTGTTGC AAGCAGCAGA TTACGCGCAG AAAAAAGGA TCTCAAGAAG ATCCTTTGAT 6060
 CTTTTCTACG GGGTCTGACG CTCAGTGGAA CGAAAACCTA CGTTAAGGGA TTTTGGTCAT 6120
 GAGATTATCA AAAAGGATCT TCACCTAGAT CCTTTTAAAT TAAAAATGAA GTTTTAAATC 6180
 AATCTAAAGT ATATATGAGT AACTTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC 6240
 ACCTATCTCA GCGATCTGTC TATTTTCGTTT ATCCATAGTT GCCTGACTCC CCGTCGTGTA 6300

GATAACTACG ATACGGGAGG GCTTACCATC TGGCCCCAGT GCTGCAATGA TACCGCGAGA 6360
CCCACGCTCA CCGGCTCCAG ATTTATCAGC AATAAACCAG CCAGCCGGAA GGGCCGAGCG 6420
CAGAAGTGGT CCTGCAACTT TATCCGCCTC CATCCAGTCT ATTAATTGTT GCCGGGAAGC 6480
TAGAGTAAGT AGTTCGCCAG TTAATAGTTT GCGCAACGTT GTTGCCATTG CTGCAGGCAT 6540
CGTGGTGTCA CGCTCGTCGT TTGGTATGGC TTCATTACAGC TCCGGTTCCC AACGATCAAG 6600
GCGAGTTACA TGATCCCCCA TGTTGTGCAA AAAAGCGGTT AGCTCCTTCG GTCCTCCGAT 6660
CGTTGTCAGA AGTAAGTTGG CCGCAGTGTT ATCACTCATG GTTATGGCAG CACTGCATAA 6720
TTCTCTTACT GTCATGCCAT CCGTAAGATG CTTTTCTGTG ACTGGTGAGT ACTCAACCAA 6780
GTCATTCTGA GAATAGTGTA TGCGGCGACC GAGTTGCTCT TGCCCGGCGT CAACACGGGA 6840
TAATACCGCG CCACATAGCA GAACTTTAAA AGTGCTCATC ATTGGAAAAC GTTCTTCGGG 6900
GCGAAAATC TCAAGGATCT TACCGCTGTT GAGATCCAGT TCGATGTAAC CCACTCGTGC 6960
ACCCAACTGA TCTTCAGCAT CTTTACTTT CACCAGCGTT TCTGGGTGAG CAAAAACAGG 7020
AAGGCAAAAT GCCGCAAAA AGGGAATAAG GGCGACACGG AAATGTTGAA TACTCATACT 7080
CTTCCTTTTT CAATATTATT GAAGCATTTA TCAGGGTTAT TGTCTCATGA GCGGATACAT 7140
ATTTGAATGT ATTTAGAAAA ATAAACAAAT AGGGGTTC CGCACATTTT CCCGAAAAGT 7200
GCCACCTGAC GTCTAAGAAA CCATTATTAT CATGACATTA ACCTATAAAA ATAGGCGTAT 7260
CACGAGGCCC TTTCGTCTTC AAGAATTAAT TCATACCAGA TCACCGAAAA CTGTCTCTCA 7320
AATGTGTCCC CCTCACACTC CCAAATTCGC GGGCTTCTGC CTCTTAGACC ACTCTACCCT 7380
ATTCCCCACA CTCACCGGAG CCAAAGCCGC GGCCCTTCCG TTTCTTTGCT 7430

<210> 12
<211> 7010
<212> DNA
<213> Artificial Sequence

<220>
<221> promoter
<222> 2806..3150
<223> BLV promoter

<220>
<221> CDS
<222> 3236..3955
<223> EYFP; enhanced yellow fluorescent protein

<220>
<221> CDS
<222> 1660..2454
<223> neomycin resistance

<220>
<221> CDS
<222> 5945..6805
<223> ampicillin resistance

<220>
<221> LTR
<222> 149..737
<223> 5' MoMuSVLTR

<220>
<221> LTR
<222> 4056..4649
<223> 3' MoMuLVLTR

<220>
<221> misc_feature
<222> 3170..3194
<223> attB1

<220>
<221> misc_feature
<222> 3980..4004
<223> attB2

<400> 12

GAATTAATTC ATACCAGATC ACCGAAAAC TGCCTCCAAA TGTGTCCCC TCACACTCCC 60
AAATTCGCGG GCTTCTGCCT CTTAGACCAC TCTACCCTAT TCCCCACACT CACCGGAGCC 120
AAAGCCGCGG CCCTTCCGTT TCTTTGCTTT TGAAAGACCC CACCCGTAGG TGGCAAGCTA 180
GCTTAAGTAA CGCCACTTTG CAAGGCATGG AAAAAATACAT AACTGAGAAT AGAAAAGTTC 240
AGATCAAGGT CAGGAACAAA GAAACAGCTG AATACCAAAC AGGATATCTG TGGTAAGCGG 300

TTCTGCCCC GGCTCAGGGC CAAGAACAGA TGAGACAGCT GAGTGATGGG CCAAACAGGA 360
TATCTGTGGT AAGCAGTTCC TGCCCCGGCT CGGGGCCAAG AACAGATGGT CCCCAGATGC 420
GGTCCAGCCC TCAGCAGTTT CTAGTGAATC ATCAGATGTT TCCAGGGTGC CCCAAGGACC 480
TGAAAATGAC CCTGTACCTT ATTTGAACTA ACCAATCAGT TCGCTTCTCG CTTCTGTTCG 540
CGCGCTTCCG CTCTCCGAGC TCAATAAAAG AGCCCACAAC CCCTCACTCG, GCGCGCCAGT 600
CTTCCGATAG ACTGCGTCGC CCGGGTACCC GTATTCCCAA TAAAGCCTCT TGCTGTTTGC 660
ATCCGAATCG TGGTCTCGCT GTTCCTTGGG AGGGTCTCCT CTGAGTGATT GACTACCCAC 720
GACGGGGGTC TTTCATTTGG GGGCTCGTCC GGGATTTGGA GACCCCTGCC CAGGGACCAC 780
CGACCCACCA CCGGGAGGTA AGCTGGCCAG CAACTTATCT GTGTCTGTCC GATTGTCTAG 840
TGTCTATGTT TGATGTTATG CGCCTGCGTC TGTACTAGTT AGCTAACTAG CTCTGTATCT 900
GGCGGACCCG TGGTGGAAC T GACGAGTTCT GAACACCCGG CCGCAACCCT GGGAGACGTC 960
CCAGGGACTT TGGGGGCCGT TTTTGTGGCC CGACCTGAGG AAGGGAGTCG ATGTGGAATC 1020
CGACCCCGTC AGGATATGTG GTTCTGGTAG GAGACGAGAA CCTAAACAG TTCCCGCCTC 1080
CGTCTGAATT TTTGCTTTCG GTTTGGAACC GAAGCCGCGC GTCTTGTCTG CTGCAGCGCT 1140
GCAGCATCGT TCTGTGTTGT CTCTGTCTGA CTGTGTTTCT GTATTTGTCT GAAAATTAGG 1200
GCCAGACTGT TACCACTCCC TTAAGTTTGA CCTTAGGTCA CTGGAAAGAT GTCGAGCGGA 1260
TCGCTCACAA CCAGTCGGTA GATGTCAAGA AGAGACGTTG GGTTACCTTC TGCTCTGCAG 1320
AATGGCCAAC CTTTAACGTC GGATGGCCGC GAGACGGCAC CTTTAACCGA GACCTCATCA 1380
CCCAGGTAA GATCAAGGTC TTTTCACCTG GCCCGCATGG ACACCCAGAC CAGGTCCCCT 1440
ACATCGTGAC CTGGGAAGCC TTGGCTTTTG ACCCCCCTCC CTGGGTCAAG CCCTTTGTAC 1500
ACCCTAAGCC TCCGCCTCCT CTTCTCCAT CCGCCCCGTC TCTCCCCCTT GAACCTCCTC 1560
GTTTCGACCC GCCTCGATCC TCCCTTTATC CAGCCCTCAC TCCTTCTCTA GGCGCCGGAA 1620
TTCCGATCTG ATCAAGAGAC AGGATGAGGA TCGTTTCGCA TGATTGAACA AGATGGATTG 1680
CACGCAGGTT CTCCGGCCGC TTGGGTGGAG AGGCTATTCG GCTATGACTG GGCACAACAG 1740
ACAATCGGCT GCTCTGATGC CGCCGTGTTT CGGCTGTCAG CGCAGGGGCG CCCGGTTCTT 1800
TTTGTCAAGA CCGACCTGTC CCGTGCCCTG AATGAACTGC AGGACGAGGC AGCGCGGCTA 1860
TCGTGGCTGG CCACGACGGG CGTTCCTTGC GCAGCTGTGC TCGACGTTGT CACTGAAGCG 1920
GGAAGGGACT GGCTGCTATT GGGCGAAGTG CCGGGGCAGG ATCTCCTGTC ATCTCACCTT 1980
GCTCCTGCCG AGAAAGTATC CATCATGGCT GATGCAATGC GGCGGCTGCA TACGCTTGAT 2040

CCGGCTACCT GCCCATTCGA CCACCAAGCG AAACATCGCA TCGAGCGAGC ACGTACTCGG 2100
 ATGGAAGCCG GTCTTGTCGA TCAGGATGAT CTGGACGAAG AGCATCAGGG GCTCGCGCCA 2160
 GCCGAAGTGT TCGCCAGGCT CAAGGCGCGC ATGCCCCGACG GCGAGGATCT CGTCGTGACC 2220
 CATGGCGATG CCTGCTTGCC GAATATCATG GTGGAAAATG GCCGCTTTTC TGGATTCATC 2280
 GACTGTGGCC GGCTGGGTGT GGC GGACCGC TATCAGGACA TAGCGTTGGC TACCCGTGAT 2340
 ATTGCTGAAG AGCTTGCGG CGAATGGGCT GACCGCTTCC TCGTGCTTTA CGGTATCGCC 2400
 GCTCCCGATT CGCAGCGCAT CGCCTTCTAT CGCCTTCTTG ACGAGTTCTT CTGAGCGGGA 2460
 CTCTGGGGTT CGAAATGACC GACCAAGCGA CGCCCAACCT GCCATCACGA GATTTTCGATT 2520
 CCACCGCCGC CTTCTATGAA AGGTTGGGCT TCGGAATCGT TTTCCGGGAC GCCGGCTGGA 2580
 TGATCCTCCA GCGCGGGGAT CTCATGCTGG AGTTCTTCGC CCACCCCGGG CTCGATCCCC 2640
 TCGCGAGTTG GTTCAGCTGC TGCCTGAGGC TGGACGACCT CGCGGAGTTC TACCGGCAGT 2700
 GCAAATCCGT CGGCATCCAG GAAACCAGCA GCGGCTATCC GCGCATCCAT GCGCCCGAAC 2760
 TGCAGGAGTG GGGAGGCACG ATGGCCGCTT TGGTCGAGGC GGATCCTAGC AGAAAAATAA 2820
 GACTTGATTG CCCCTTAAAA TTACAACGTC TAGAAAAATGA ATGGCTCTCC CGCCTTTTTT 2880
 GAGGGGGAAT CATTTGTATG AAAGATCATG CCGACCTAGG CGCCGCCACC GCGCCGTAAA 2940
 CCAGACAGAG ACGTCAGCTG CCAGAAAAGC TGGTGACGGC AGCTGGTGGC TAGAATCCCC 3000
 GTACCTCCCC AACTTCCCCCT TTCCCGAAAA ATCCACACCC TGAGCTGCTG ACCTCACCTG 3060
 CTGATAAATT AATAAAATGC CGGCCCTGTC GAGTTAGCGG CACCAGAAGC GTTCTTCTCC 3120
 TGAGACCCCTC GTGCTCAGCT CTCGGTCTCG CCTCGAGAAG CTTGTTATCA CAAGTTTGTA 3180
 CAAAAAAGCA GGCTTCGAAG GAGATAGAAC CAATTCTCTA AGGAAATACT TAACC 3235
 ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC 3280
 Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile
 1 5 10 15
 CTG GTC GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC GTG 3325
 Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val
 20 25 30
 TCC GGC GAG GGC GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG 3370
 Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu
 35 40 45
 AAG TTC ATC TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC 3415
 Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr
 50 55 60

CTC GTG ACC ACC TTC GGC TAC GGC CTG CAG TGC TTC GCC CGC TAC Leu Val Thr Thr Phe Gly Tyr Gly Leu Gln Cys Phe Ala Arg Tyr 65 70 75	3460
CCC GAC CAC ATG AAG CAG CAC GAC TTC TTC AAG TCC GCC ATG CCC Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro 80 85 90	3505
GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC AAG GAC GAC GGC Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly 95 100 105	3550
AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC CTG Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu 110 115 120	3595
GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly 125 130 135	3640
AAC ATC CTG GGG CAC AAG CTG GAG TAC AAC TAC AAC AGC CAC AAC Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn 140 145 150	3685
GTC TAT ATC ATG GCC GAC AAG CAG AAG AAC GGC ATC AAG GTG AAC Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn 155 160 165	3730
TTC AAG ATC CGC CAC AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala 170 175 180	3775
GAC CAC TAC CAG CAG AAC ACC CCC ATC GGC GAC GGC CCC GTG CTG Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu 185 190 195	3820
CTG CCC GAC AAC CAC TAC CTG AGC TAC CAG TCC GCC CTG AGC AAA Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Ala Leu Ser Lys 200 205 210	3865
GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG CTG GAG TTC GTG Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val 215 220 225	3910
ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC AAG TAA Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys *** 230 235	3955
AGCGG	3960
CCGCACTCGA GATATCTAGA CCCAGCTTTC TTGTACAAAG TGGTGATAAC ATCGATAAAA	4020
TAAAAGATTT TATTTAGTCT CCAGAAAAAG GGGGGAATGA AAGACCCAC CTGTAGGTTT	4080
GGCAAGCTAG CTTAAGTAAC GCCATTTTGC AAGGCATGGA AAAATACATA ACTGAGAATA	4140

GAGAAGTTCA GATCAAGGTC AGGAACAGAT GGAACAGCTG AATATGGGCC AAACAGGATA 4200
TCTGTGGTAA GCAGTTCCTG CCCC GGCTCA GGGCCAAGAA CAGATGGAAC AGCTGAATAT 4260
GGGCCAAACA GGATATCTGT GGTAAGCAGT TCCTGCCCCG GCTCAGGGCC AAGAACAGAT 4320
GGTCCCCAGA TGCGGTCCAG CCCTCAGCAG TTTCTAGAGA ACCATCAGAT GTTTCAGGG 4380
TGCCCCAAGG ACCTGAAATG ACCCTGTGCC TTATTTGAAC TAACCAATCA GTTCGCTTCT 4440
CGCTTCTGTT CGCGCGCTTC TGCTCCCCGA GCTCAATAAA AGAGCCCACA ACCCCTCACT 4500
CGGGGCGCCA GTCCTCCGAT TGA CTGAGTC GCCCGGGTAC CCGTGTATCC AATAAACCT 4560
CTTG CAGTTG CATCCGACTT GTGGTCTCGC TGTTCTTGG GAGGGTCTCC TCTGAGTGAT 4620
TGA CTACCCG TCAGCGGGG TCTTTTATTT GGGGGCTCGT CCGGGATCGG GAGACCCCTG 4680
CCCAGGGACC ACCGACCCAC CACCGGGAGG TAAGCTGGCT GCCTCGCGCG TTTCGGTGAT 4740
GACGGTGAAA ACCTCTGACA CATGCAGCTC CCGGAGACGG TCACAGCTTG TCTGTAAGCG 4800
GATGCCGGGA GCAGACAAGC CCGTCAGGGC GCGTCAGCGG GTGTTGGCGG GTGTCGGGGC 4860
GCAGCCATGA CCCAGTCACG TAGCGATAGC GGAGTGTATA CTGGCTTAAC TATGCGGCAT 4920
CAGAGCAGAT TGTA CTGAGA GTGCACCATA TGCGGTGTGA AATACCGCAC AGATGCGTAA 4980
GGAGAAAATA CCGCATCAGG CGCTCTTCCG CTCCTCGCT CACTGACTCG CTGCGCTCGG 5040
TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG TTATCCACAG 5100
AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAG CCAGCAAAAG GCCAGGAACC 5160
GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC ATAGGCTCCG CCCCCCTGAC GAGCATCACA 5220
AAAATCGACG CTCAAGTCAG AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT 5280
TTCCCCCTGG AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC 5340
TGTCCGCTT TCTCCCTTCG GGAAGCGTGG CGCTTTCTCA TAGCTCACGC TG TAGGTATC 5400
TCAGTTCGGT GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT GCACGAACCC CCCGTT CAGC 5460
CCGACCGCTG CGCCTTATCC GGTA ACTATC GTCTTGAGTC CAACCCGGTA AGACACGACT 5520
TATCGCCACT GGCAGCAGCC ACTGGTAACA GGATTAGCAG AGCGAGGTAT GTAGGCGGTG 5580
CTACAGAGTT CTTGAAGTGG TGGCCTAACT ACGGCTACAC TAGAAGGACA GTATTTGGTA 5640
TCTGCGCTCT GCTGAAGCCA GTTACCTTCG GAAAAAGAGT TGGTAGCTCT TGATCCGGCA 5700
AACAAACCAC CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA GCAGCAGATT ACGCGCAGAA 5760
AAAAAGGATC TCAAGAAGAT CCTTTGATCT TTTCTACGGG GTCTGACGCT CAGTGGAACG 5820
AAAACTCAG TTAAGGGATT TTGGTCATGA GATTATCAA AAGGATCTTC ACCTAGATCC 5880

TTTTAAATTA AAAATGAAGT TTTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG 5940
 ACAGTTACCA ATGCTTAATC AGTGAGGCAC CTATCTCAGC GATCTGTCTA TTTCGTTTCAT 6000
 CCATAGTTGC CTGACTCCCC GTCGTGTAGA TAACTACGAT ACGGGAGGGC TTACCATCTG 6060
 GCCCCAGTGC TGCAATGATA CCGCGAGACC CACGCTCACC GGCTCCAGAT TTATCAGCAA 6120
 TAAACCAGCC AGCCGGAAGG GCCGAGCGCA GAAGTGGTCC TGCAACTTTA TCCGCCTCCA 6180
 TCCAGTCTAT TAATTGTTGC CGGGAAGCTA GAGTAAGTAG TTCGCCAGTT AATAGTTTGC 6240
 GCAACGTTGT TGCCATTGCT GCAGGCATCG TGGTGTACAG CTCGTCGTTT GGTATGGCTT 6300
 CATTCAGCTC CGGTTCCTCA CGATCAAGGC GAGTTACATG ATCCCCCATG TTGTGCAAAA 6360
 AAGCGGTTAG CTCCTTCGGT CCTCCGATCG TTGTGAGAAG TAAGTTGGCC GCAGTGTAT 6420
 CACTCATGGT TATGGCAGCA CTGCATAATT CTCTTACTGT CATGCCATCC GTAAGATGCT 6480
 TTTCTGTGAC TGGTGAGTAC TCAACCAAGT CATTCTGAGA ATAGTGTATG CGGCGACCGA 6540
 GTTGCTCTTG CCCGGCGTCA ACACGGGATA ATACCGCGCC ACATAGCAGA ACTTTAAAAG 6600
 TGCTCATCAT TGGAAAACGT TCTTCGGGGC GAAACTCTC AAGGATCTTA CCGCTGTTGA 6660
 GATCCAGTTC GATGTAACCC ACTCGTGCAC CCAACTGATC TTCAGCATCT TTTACTTTCA 6720
 CCAGCGTTTC TGGGTGAGCA AAAACAGGAA GGCAAAATGC CGCAAAAAG GGAATAAGGG 6780
 CGACACGGAA ATGTTGAATA CTCATACTCT TCCTTTTTCA ATATTATTGA AGCATTTATC 6840
 AGGGTTATTG TCTCATGAGC GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG 6900
 GGGTTCCGCG CACATTTCCC CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA 6960
 TGACATTAAC CTATAAAAAT AGGCGTATCA CGAGGCCCTT TCGTCTTCAA 7010

<210> 13
<211> 7121
<212> DNA
<213> Artificial Sequence

<220>
<221> promoter
<222> 2806..3261
<223> HIV promoter

<220>
<221> CDS
<222> 3347..4066
<223> EYFP; enhanced yellow fluorescent protein

<220>
<221> CDS
<222> 1660..2454
<223> neomycin resistance

<220>
<221> CDS
<222> 6056..6916
<223> ampicillin resistance

<220>
<221> LTR
<222> 149..737
<223> 5' MoMuSVLTR

<220>
<221> LTR
<222> 4167..4760
<223> 3' MoMuLVLTR

<220>
<221> misc_feature
<222> 3281..3305
<223> attB1

<220>
<221> misc_feature
<222> 4091..4115
<223> attB2

<400> 13

GAATTAATTC ATACCAGATC ACCGAAAAC TGCCTCCAAA TGTGTCCCC TCACACTCCC 60
AAATTCGCGG GCTTCTGCCT CTTAGACCAC TCTACCCTAT TCCCCACACT CACCGGAGCC 120
AAAGCCGCGG CCCTTCCGTT TCTTTGCTTT TGAAAGACCC CACCCGTAGG TGGCAAGCTA 180
GCTTAAGTAA CGCCACTTTG CAAGGCATGG AAAAATACAT AACTGAGAAT AGAAAAGTTC 240
AGATCAAGGT CAGGAACAAA GAAACAGCTG AATACCAAAC AGGATATCTG TGGTAAGCGG 300

TTCTGCCCC GGCTCAGGGC CAAGAACAGA TGAGACAGCT GAGTGATGGG CCAAACAGGA 360
 TATCTGTGGT AAGCAGTTCC TGCCCCGGCT CGGGGCCAAG AACAGATGGT CCCCAGATGC 420
 GGTCCAGCCC TCAGCAGTTT CTAGTGAATC ATCAGATGTT TCCAGGGTGC CCCAAGGACC 480
 TGAAAATGAC CCTGTACCTT ATTTGAACTA ACCAATCAGT TCGCTTCTCG CTTCTGTTCG 540
 CGCGCTTCCG CTCTCCGAGC TCAATAAAAG AGCCCACAAC CCCTCACTCG GCGCGCCAGT 600
 CTTCCGATAG ACTGCGTCGC CCGGGTACCC GTATTCCCAA TAAAGCCTCT TGCTGTTTGC 660
 ATCCGAATCG TGGTCTCGCT GTTCCTTGGG AGGGTCTCCT CTGAGTGATT GACTACCCAC 720
 GACGGGGGTC TTTCATTTGG GGGCTCGTCC GGGATTTGGA GACCCCTGCC CAGGGACCAC 780
 CGACCCACCA CCGGGAGGTA AGCTGGCCAG CAACTTATCT GTGTCTGTCC GATTGTCTAG 840
 TGTCTATGTT TGATGTTATG CGCCTGCGTC TGTACTAGTT AGCTAACTAG CTCTGTATCT 900
 GGC GGACCCG TGGTGGAACT GACGAGTTCT GAACACCCGG CCGCAACCTT GGGAGACGTC 960
 CCAGGGACTT TGGGGGCCGT TTTTGTGGCC CGACCTGAGG AAGGGAGTCG ATGTGGAATC 1020
 CGACCCCGTC AGGATATGTG GTTCTGGTAG GAGACGAGAA CCTAAACAG TTCCCGCCTC 1080
 CGTCTGAATT TTTGCTTTTCG GTTTGGAACC GAAGCCGCGC GTCTTGTCTG CTGCAGCGCT 1140
 GCAGCATCGT TCTGTGTTGT CTCTGTCTGA CTGTGTTTCT GTATTTGTCT GAAAATTAGG 1200
 GCCAGACTGT TACCACTCCC TTAAGTTTGA CCTTAGGTCA CTGGAAGAT GTCGAGCGGA 1260
 TCGCTCACAA CCAGTCGGTA GATGTCAAGA AGAGACGTTG GGTTACCTTC TGCTCTGCAG 1320
 AATGGCCAAC CTTTAACGTC GGATGGCCGC GAGACGGCAC CTTTAACCGA GACCTCATCA 1380
 CCCAGGTAA GATCAAGGTC TTTTCACCTG GCCCGCATGG ACACCCAGAC CAGGTCCCCT 1440
 ACATCGTGAC CTGGGAAGCC TTGGCTTTTG ACCCCCCTCC CTGGGTCAAG CCCTTTGTAC 1500
 ACCCTAAGCC TCCGCCTCCT CTTCCCTCCAT CCGCCCCGTC TCTCCCCCTT GAACCTCCTC 1560
 GTTCGACCCC GCCTCGATCC TCCCTTTATC CAGCCCTCAC TCCTTCTCTA GGC GCCGGA 1620
 TTCCGATCTG ATCAAGAGAC AGGATGAGGA TCGTTTCGCA TGATTGAACA AGATGGATTG 1680
 CACGCAGGTT CTCCGGCCGC TTGGGTGGAG AGGCTATTCG GCTATGACTG GGCACAACAG 1740
 ACAATCGGCT GCTCTGATGC CGCCGTGTTT CGGCTGTCAG CGCAGGGGCG CCCGGTTCTT 1800
 TTTGTCAAGA CCGACCTGTC CGGTGCCCTG AATGAACTGC AGGACGAGGC AGCGCGGCTA 1860
 TCGTGGCTGG CCACGACGGG CGTTCCTTGC GCAGCTGTGC TCGACGTTGT CACTGAAGCG 1920
 GGAAGGGACT GGCTGCTATT GGGCGAAGTG CCGGGGCAGG ATCTCCTGTC ATCTCACCTT 1980

GCTCCTGCCG AGAAAGTATC CATCATGGCT GATGCAATGC GGCGGCTGCA TACGCTTGAT 2040
CCGGCTACCT GCCCATTCGA CCACCAAGCG AAACATCGCA TCGAGCGAGC ACGTACTCGG 2100
ATGGAAGCCG GTCTTGTCGA TCAGGATGAT CTGGACGAAG AGCATCAGGG GCTCGCGCCA 2160
GCCGAACTGT TCGCCAGGCT CAAGGCGCGC ATGCCCCGACG GCGAGGATCT CGTCGTGACC 2220
CATGGCGATG CCTGCTTGCC GAATATCATG GTGGAAAATG GCCGCTTTTC TGGATTCATC 2280
GACTGTGGCC GGCTGGGTGT GGCGGACCGC TATCAGGACA TAGCGTTGGC TACCCGTGAT 2340
ATTGCTGAAG AGCTTGCGG CGAATGGGCT GACCGCTTCC TCGTGCTTTA CGGTATCGCC 2400
GCTCCCGATT CGCAGCGCAT CGCCTTCTAT CGCCTTCTTG ACGAGTTCTT CTGAGCGGGA 2460
CTCTGGGGTT CGAAATGACC GACCAAGCGA CGCCCAACCT GCCATCACGA GATTTCGATT 2520
CCACCGCCGC CTTCTATGAA AGGTTGGGCT TCGGAATCGT TTTCCGGGAC GCCGGCTGGA 2580
TGATCCTCCA GCGCGGGGAT CTCATGCTGG AGTTCTTCGC CCACCCCGGG CTCGATCCCC 2640
TCGCGAGTTG GTTCAGCTGC TGCCTGAGGC TGGACGACCT CGCGGAGTTC TACCGGCAGT 2700
GCAAATCCGT CGGCATCCAG GAAACCAGCA GCGCTATCC GCGCATCCAT GCGCCGAAC 2760
TGCAGGAGTG GGGAGGCACG ATGGCCGCTT TGGTCGAGGC GGATCCTGGA AGGGCTAATT 2820
TGGTCCCAA GAAGACAAGA GATCCTTGAT CTGTGGATCT ACCACACACA AGGCTACTTC 2880
CCTGATTGGC AGAATTACAC ACCAGGGCCA GGGATCAGAT ATCCACTGAC CTTTGGATGG 2940
TGCTTCAAGC TAGTACCAGT TGAGCCAGAG AAGGTAGAAG AGGCCAATGA AGGAGAGAAC 3000
AACAGCTTGT TACACCCTAT GAGCCTGCAT GGGATGGAGG ACGCGGAGAA AGAAGTGTTA 3060
GTGTGGAGGT TTGACAGCAA ACTAGCATTT CATCACATGG CCCGAGAGCT GCATCCGGAG 3120
TACTACAAAG ACTGCTGACA TCGAGCTTTC TACAAGGGAC TTTCCGCTGG GGACTTTCCA 3180
GGGAGGCGTG GCCTGGGCGG GACTGGGGAG TGGCGTCCCT CAGATGCTGC ATATAAGCAG 3240
CTGCTTTTTG CCTGTACTGG GCCTCGAGAA GCTTGTATC ACAAGTTTGT AAAAAAAGC 3300
AGGCTTCGAA GGAGATAGAA CCAATTCTCT AAGGAAATAC TTAACC 3346
ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC 3391
Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile
1 5 10 15
CTG GTC GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC GTG 3436
Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val
20 25 30
TCC GGC GAG GGC GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG 3481
Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu
35 40 45

AAG TTC ATC TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr 50 55 60	3526
CTC GTG ACC ACC TTC GGC TAC GGC CTG CAG TGC TTC GCC CGC TAC Leu Val Thr Thr Phe Gly Tyr Gly Leu Gln Cys Phe Ala Arg Tyr 65 70 75	3571
CCC GAC CAC ATG AAG CAG CAC GAC TTC TTC AAG TCC GCC ATG CCC Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro 80 85 90	3616
GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC AAG GAC GAC GGC Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly 95 100 105	3661
AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC CTG Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu 110 115 120	3706
GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly 125 130 135	3751
AAC ATC CTG GGG CAC AAG CTG GAG TAC AAC TAC AAC AGC CAC AAC Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn 140 145 150	3796
GTC TAT ATC ATG GCC GAC AAG CAG AAG AAC GGC ATC AAG GTG AAC Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn 155 160 165	3841
TTC AAG ATC CGC CAC AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala 170 175 180	3886
GAC CAC TAC CAG CAG AAC ACC CCC ATC GGC GAC GGC CCC GTG CTG Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu 185 190 195	3931
CTG CCC GAC AAC CAC TAC CTG AGC TAC CAG TCC GCC CTG AGC AAA Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Ala Leu Ser Lys 200 205 210	3976
GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG CTG GAG TTC GTG Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val 215 220 225	4021
ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC AAG TAA Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys *** 230 235	4066
AGCGGCCGCA CTCGAGATAT CTAGACCCAG CTTTCTTGTA CAAAGTGGTG ATAACATCGA 4126	
TAAATAAAAA GATTTTATTT AGTCTCCAGA AAAAGGGGGG AATGAAAGAC CCCACCTGTA 4186	

GGTTTGGCAA GCTAGCTTAA GTAACGCCAT TTTGCAAGGC ATGGAAAAAT ACATAACTGA 4246
GAATAGAGAA GTTCAGATCA AGGTCAGGAA CAGATGGAAC AGCTGAATAT GGGCCAAACA 4306
GGATATCTGT GGTAAGCAGT TCCTGCCCCG GCTCAGGGCC AAGAACAGAT GGAACAGCTG 4366
AATATGGGCC AAACAGGATA TCTGTGGTAA GCAGTTCCTG CCCC GGCTCA GGGCCAAGAA 4426
CAGATGGTCC CCAGATGCGG TCCAGCCCTC AGCAGTTTCT AGAGAACCAT CAGATGTTTC 4486
CAGGGTGCCC CAAGGACCTG AAATGACCCT GTGCCTTATT TGAAC TAACC AATCAGTTCG 4546
CTTCTCGCTT CTGTTGCGGC GCTTCTGCTC CCCGAGCTCA ATAAAAGAGC CCACAACCCC 4606
TCACTCGGGG CGCCAGTCCT CCGATTGACT GAGTCGCCCC GGTACCCGTG TATCCAATAA 4666
ACCCCTCTTG AGTTGCATCC GACTTGTGGT CTCGCTGTTC CTTGGGAGGG TCTCCTCTGA 4726
GTGATTGACT ACCCGTCAGC GGGGGTCTTT CATTTGGGGG CTCGTCCGGG ATCGGGAGAC 4786
CCCTGCCCAG GGACCACCGA CCCACCACCG GGAGGTAAGC TGGCTGCCTC GCGCGTTTCG 4846
GTGATGACGG TGAAAACCTC TGACACATGC AGCTCCCGGA GACGGTCACA GCTTGTCTGT 4906
AAGCGGATGC CGGGAGCAGA CAAGCCCGTC AGGGCGCGTC AGCGGGTGTT GGCGGGTGTC 4966
GGGGCGCAGC CATGACCCAG TCACGTAGCG ATAGCGGAGT GTATACTGGC TTAACATATGC 5026
GGCATCAGAG CAGATTGTAC TGAGAGTGCA CCATATGCGG TGTGAAATAC CGCACAGATG 5086
CGTAAGGAGA AAATACCGCA TCAGGCGCTC TTCCGCTTCC TCGCTCACTG ACTCGCTGCG 5146
CTCGGTCGTT CGGCTGCGGC GAGCGGTATC AGCTCACTCA AAGGCGGTAA TACGGTTATC 5206
CACAGAATCA GGGGATAACG CAGGAAAGAA CATGTGAGCA AAAGGCCAGC AAAAGGCCAG 5266
GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCATAGG CTCCGCCCCC CTGACGAGCA 5326
TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG ACAGGACTAT AAAGATACCA 5386
GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT CCGACCCTGC CGCTTACCGG 5446
ATACCTGTCC GCCTTTCTCC CTTGCGGAAG CGTGGCGCTT TCTCATAGCT CACGCTGTAG 5506
GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC TGTGTGCACG AACCCCCCGT 5566
TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT GAGTCCAACC CGGTAAGACA 5626
CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT AGCAGAGCGA GGTATGTAGG 5686
CGGTGCTACA GAGTTCCTGA AGTGGTGGCC TAACTACGGC TACACTAGAA GGACAGTATT 5746
TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTTGCGAAAA AGAGTTGGTA GCTCTTGATC 5806
CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT 'TGCAAGCAGC AGATTACGCG 5866
CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT ACGGGGTCTG ACGCTCAGTG 5926

GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA TCAAAAAGGA TCTTCACCTA 5986
GATCCTTTTA AATTAAAAAT GAAGTTTAA ATCAATCTAA AGTATATATG AGTAAACTTG 6046
GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC TCAGCGATCT GTCTATTTTCG 6106
TTCATCCATA GTTGCCCTGAC TCCCCGTCGT GTAGATAACT ACGATACGGG AGGGCTTACC 6166
ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC TCACCGGCTC CAGATTTATC 6226
AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT GGTCTTGCAA CTTTATCCGC 6286
CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA AGTAGTTCGC CAGTTAATAG 6346
TTTGCGCAAC GTTGTGCCA TTGCTGCAGG CATCGTGGTG TCACGCTCGT CGTTTGGTAT 6406
GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCGAGTT ACATGATCCC CCATGTTGTG 6466
CAAAAAGCG GTTAGTCTCT TCGGTCCTCC GATCGTTGTC AGAAGTAAGT TGGCCGAGT 6526
GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT ACTGTCATGC CATCCGTAAG 6586
ATGCTTTTCT GTGACTGGTG AGTACTCAAC CAAGTCATTC TGAGAATAGT GTATGCGGCG 6646
ACCGAGTTGC TCTTGCCCGG CGTCAACACG GGATAATACC GCGCCACATA GCAGAACTTT 6706
AAAAGTGCTC ATCATTGGA AACGTTCTTC GGGGCGAAAA CTCTCAAGGA TCTTACCGCT 6766
GTTGAGATCC AGTTCGATGT AACCCACTCG TGCACCCAAC TGATCTTCAG CATCTTTTAC 6826
TTTACCAGC GTTTCTGGGT GAGCAAAAAC AGGAAGGCAA AATGCCGCAA AAAAGGGAAT 6886
AAGGGCGACA CGGAAATGTT GAATACTCAT ACTCTTCCTT TTTCAATATT ATTGAAGCAT 6946
TTATCAGGGT TATTGTCTCA TGAGCGGATA CATATTTGAA TGTATTTAGA AAAATAAACA 7006
AATAGGGGTT CCGCGCACAT TTCCCCGAAA AGTGCCACCT GACGTCTAAG AAACCATTAT 7066
TATCATGACA TTAACCTATA AAAATAGGCG TATCACGAGG CCCTTTCGTC TTCAA 7121